

HỘI THẢO: DỮ LIỆU NGHIÊN CỨU KHOA HỌC MỞ VÀ LƯU TRỮ DỮ LIỆU NGHIÊN CỨU KHOA HỌC MỞ DO TRUNG TÂM CNTT - BỘ KHCN TỔ CHỨC TẠI TRƯỜNG ĐH KINH TẾ TP. HỒ CHÍ MINH, 13/09/2023

DỮ LIỆU NGHIÊN CỨU MỞ & CÁC NGUYÊN TẮC DỮ LIỆU FAIR

LÊ TRUNG NGHĨA

VIỆN NGHIÊN CỨU, ĐÀO TẠO VÀ PHÁT TRIỂN TÀI NGUYÊN GIÁO DỤC MỞ,
HIỆP HỘI CÁC TRƯỜNG ĐẠI HỌC, CAO ĐẲNG VIỆT NAM

Email: nghialt@inoer.vn; letrungnghia.foss@gmail.com;

Blogs: <http://vnfoss.blogspot.com/>

<http://letrungnghia.mangvn.org/>

1. Đặt vấn đề
2. Dữ liệu nghiên cứu mở là gì?
3. FAIR là gì
4. Quan hệ giữa dữ liệu FAIR và dữ liệu (nghiên cứu) mở
5. Kết luận và gợi ý



Tài liệu này mang giấy phép [Creative Commons Attribution v4.0](http://creativecommons.org/licenses/by/4.0/).
Bạn được trao các quyền để sử dụng, chia sẻ, sao chép, phân phối, phân phối lại, áp dụng, pha trộn, tùy biến và xây dựng dựa trên các tư liệu của nó, kể cả thương mại hóa, miễn là bạn ghi nhận công (các) tác giả gốc ban đầu của tài liệu.
Một bản sao giấy phép này có tại: <http://creativecommons.org/licenses/by/4.0/>.



Đặt vấn đề



- Ngày 23/11/2021, **Khuyến nghị Khoa học Mở** của **UNESCO** đã được 193 quốc gia thành viên thông qua.

“Khuyến nghị này đưa ra định nghĩa chung, các giá trị, các nguyên tắc và các tiêu chuẩn được chia sẻ cho khoa học mở ở mức quốc tế và đề xuất một tập hợp các hành động có lợi cho việc vận hành khoa học mở một cách công bằng và bình đẳng cho tất cả mọi người ở các mức cá nhân, cơ sở, quốc gia, khu vực và quốc tế” (Xem mục 2 của Phần I. Mục tiêu và các mục đích của Khuyến nghị).

- Trong số các nguyên tắc và tiêu chuẩn được chia sẻ cho khoa học mở ở mức quốc tế được nhắc tới ở trên, có **các nguyên tắc dữ liệu FAIR** dành cho việc **quản lý tốt các dữ liệu nghiên cứu mở**.



Dữ liệu nghiên cứu mở là gì? -1

- Định nghĩa Kiến thức khoa học mở (KT KHM)

“Kiến thức khoa học mở tham chiếu tới việc **truy cập mở** tới các xuất bản phẩm khoa học, **dữ liệu nghiên cứu**, **siêu dữ liệu**, tài nguyên giáo dục mở, phần mềm và mã nguồn và phần cứng mà chúng **sẵn sàng trong phạm vi công cộng** hoặc **có bản quyền** và **được cấp phép theo một giấy phép mở** cho phép truy cập, sử dụng lại, tái mục đích, tùy chỉnh và phân phối theo các điều kiện nhất định, được cung cấp cho **tất cả các tác nhân ngay lập tức hoặc nhanh nhất có thể**,... và **miễn phí**.”

- Định nghĩa Dữ liệu nghiên cứu mở

... **tuân thủ thừa nhận ghi công**. Dữ liệu nghiên cứu mở là sẵn sàng ở định dạng kịp thời và thân thiện với người sử dụng, người và **máy đọc được và hành động được**, phù hợp với các nguyên tắc của điều hành và quản trị dữ liệu tốt, ấy là các nguyên tắc Tìm thấy được, Truy cập được, Tương hợp được, Sử dụng lại được – **FAIR** (Findable, Accessible, Interoperable, Reusable), được sự giám tuyển và duy trì thường xuyên hỗ trợ.



Hình do Lê Trung Nghĩa tùy chỉnh từ tài liệu Khuyến nghị Khoa học Mở của UNESCO



Dữ liệu nghiên cứu mở là gì? -2

1. Dữ liệu nghiên cứu mở là một thành phần của KT KHM

a) Sẵn sàng trong phạm vi công cộng (PVCC): - Giải thích về tác phẩm nằm trong PVCC; - Luật SHTT Việt Nam về tác phẩm nằm trong PVCC; - Truy cập tới tác phẩm nằm trong PVCC thế nào?; - Số hóa tác phẩm nằm trong PVCC thế nào?

b) **Có bản quyền và được cấp phép theo một giấy phép mở** cho phép truy cập, sử dụng lại, tái mục đích, tùy chỉnh và phân phối theo các điều kiện nhất định, được cung cấp cho **tất cả các tác nhân** ngay lập tức hoặc nhanh nhất có thể,... và **miễn phí** → **Không phải giấy phép mở bất kỳ!**

3 “Lớp” giấy phép

2. Dữ liệu mở gồm: dữ liệu số và tuần tự; thô và đã qua xử lý; & siêu dữ liệu đi kèm.

a) Người sử dụng phải tuân thủ **thừa nhận ghi công đúng cách**. Ngược lại là ‘ăn cắp’!

b) Cả người và máy đọc được.

CC REL (CC Right Expression Language)


→ Phân biệt người hay là máy đọc được?

c) Rất quan trọng: Tuân thủ **FAIR**

1. Legal Code: cho luật sư

2. Commons Deed - Chứng thư chung

3. Machine Readable: **Máy đọc được!**

 This work is licensed under a Creative Commons Attribution 4.0 International License .	<code>
This work is licensed under a Creative Commons Attribution 4.0 International License.</code>
Cấp phép mở CC Người đọc được!	Cấp phép mở CC theo Ngôn ngữ Thể hiện Quyền - REL (Rights Expression Language): Máy đọc được!

Dữ liệu phải đảm bảo cả người và máy đều đọc được → máy xử lý tự động → **FAIR**

Các nguyên tắc dữ liệu FAIR

Theo FORCE11: <https://force11.org/info/the-fair-data-principles/>, truy cập 21/07/2023

Để Tìm thấy được - **F**indable:

- F1. Dữ liệu/siêu dữ liệu được gắn một mã nhận diện thường trực duy nhất toàn cầu và vĩnh viễn.
- F2. Dữ liệu được mô tả với siêu dữ liệu phong phú.
- F3. Dữ liệu/siêu dữ liệu được đăng ký hoặc lập chỉ mục trong một tài nguyên tìm kiếm được.
- F4. Siêu dữ liệu chỉ định một mã nhận diện dữ liệu.

Để Truy cập được - **A**ccessible:

- A1. Dữ liệu/siêu dữ liệu là truy xuất được bằng mã nhận diện của chúng bằng việc sử dụng một giao thức truyền thông được tiêu chuẩn hóa.
 - A1.1 Giao thức đó là mở, tự do không mất tiền, và triển khai được vạn năng.
 - A1.2 Giao thức đó cho phép thủ tục xác thực và ủy quyền bất cứ ở đâu cần thiết.
- A2. Siêu dữ liệu là truy cập được, ngay cả khi dữ liệu không còn tồn tại.

Để tương hợp được - **I**nteroperable:

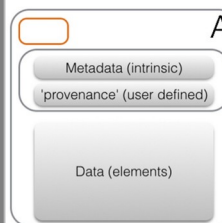
- I1. Dữ liệu/siêu dữ liệu sử dụng ngôn ngữ chính thống, truy cập được, được chia sẻ, và áp dụng được rộng rãi để trình bày kiến thức.
- I2. Dữ liệu/siêu dữ liệu sử dụng các từ vựng tuân thủ các nguyên tắc FAIR.
- I3. Dữ liệu/siêu dữ liệu bao gồm các tham chiếu đủ điều kiện tới dữ liệu/siêu dữ liệu khác.

Để Sử dụng lại được - **R**eusable:

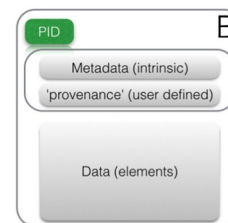
- R1. Dữ liệu/siêu dữ liệu có nhiều thuộc tính chính xác và có liên quan.
 - R1.1 Dữ liệu/siêu dữ liệu được phát hành với một giấy phép sử dụng dữ liệu rõ ràng và truy cập được.
 - R1.2 Dữ liệu/siêu dữ liệu có liên kết với nguồn gốc xuất xứ của chúng.
 - R1.3 Dữ liệu/siêu dữ liệu đáp ứng các tiêu chuẩn cộng đồng của lĩnh vực có liên quan.

Data as increasingly FAIR Digital Objects

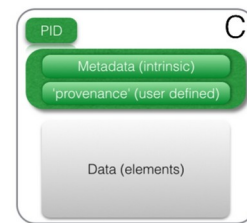
Re-useless data (80%)



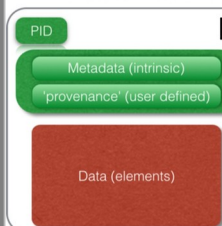
Findable



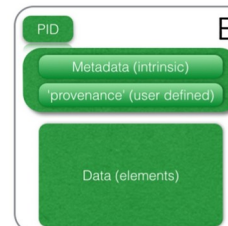
FAIR metadata



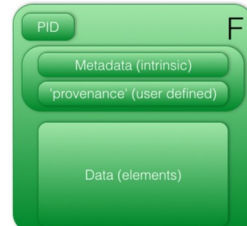
FAIR data-
restricted access



FAIR data-
Open Access



FAIR data-
Open Access/Functionally Linked

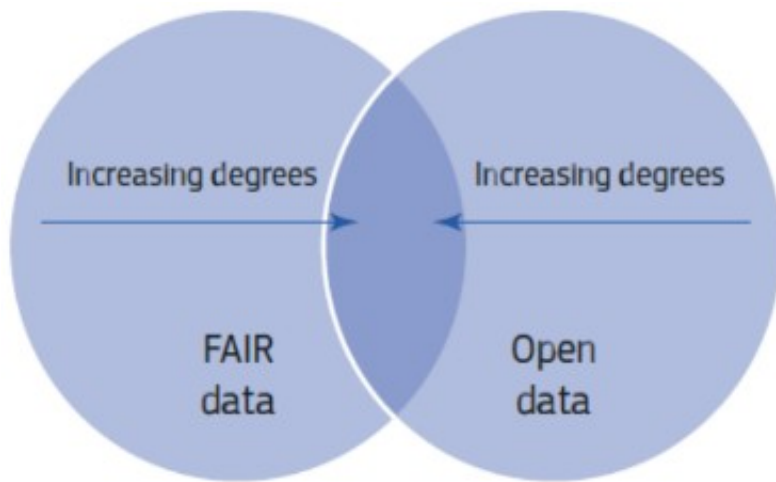


Persistent Identifier (PID)

Không có PID, dữ liệu hầu như không sử dụng lại được, điều hoàn toàn trái ngược với FAIR! Để (siêu) dữ liệu là FAIR, phải có PID (F1, A1) và giấy phép sử dụng rõ ràng (R 1.1) - **tối thiểu!**

“Dữ liệu có thể là FAIR hoặc Mở, vừa là FAIR vừa là Mở, hoặc không là FAIR không là Mở. **Các lợi ích lớn nhất khi dữ liệu vừa là FAIR vừa là Mở**, vì khi không có các hạn chế sẽ hỗ trợ được rộng lớn nhất có thể cho sử dụng lại, và sử dụng lại ở phạm vi rộng. Để tối đa hóa những lợi ích của việc biến dữ liệu FAIR thành hiện thực, và **trong ngữ cảnh của các sáng kiến Khoa học Mở**, các nguyên tắc FAIR nên được triển khai kết hợp với yêu cầu chính sách rằng **dữ liệu nghiên cứu nên là Mở mặc định** - đó là, Mở trừ phi có lý do tốt để hạn chế truy cập hoặc sử dụng lại”

EC (2018): *Turning FAIR into Reality*. Bản tiếng Việt, tr. 35



“Dữ liệu sẽ không thực sự sử dụng lại được trừ phi nó là **MỞ**, nghĩa là, **sẵn sàng theo một giấy phép mở** và với các chi phí tượng trưng (**trong hầu hết các trường hợp chi phí bằng 0**), và **tính mở thường đi cùng với sự triển khai các nguyên tắc FAIR**”

EC (2018): *Cost of Not having FAIR research data*.

Bản tiếng Việt, tr. 14

- Để dữ liệu nghiên cứu mở tuân thủ FAIR không dễ!
- Nhà nghiên cứu cần có năng lực cần thiết: (1) cấp phép mở; (2) phân biệt giữa dữ liệu mở và dữ liệu không mở, giữa dữ liệu nằm trong PVCC và được cấp phép mở đúng → các năng lực khoa học mở → **các năng lực tài nguyên giáo dục mở (TNGDM)**.
- Một trong các cách thức tham gia khoa học mở là: Nhà nghiên cứu cần có các kỹ năng khoa học mở cốt lõi như được **NASA gợi ý**, bao gồm khả năng làm việc và chỉ định các PID, ví dụ như DOI, ORCID, .v.v.

Type	Type
<input type="checkbox"/> Publication (1816828) - <ul style="list-style-type: none"><input type="checkbox"/> Article (1152581)<input type="checkbox"/> Taxonomictreatment (471803)<input type="checkbox"/> Conferencepaper (81474)<input type="checkbox"/> Report (20274)<input type="checkbox"/> Other (18921)<input type="checkbox"/> Book (14820)<input type="checkbox"/> Deliverable (14422)<input type="checkbox"/> Section (13492)<input type="checkbox"/> Thesis (9008)<input type="checkbox"/> Preprint (7769)<input type="checkbox"/> Workingpaper (5931)<input type="checkbox"/> Technicalnote (2825)<input type="checkbox"/> Proposal (784)<input type="checkbox"/> Softwaredocumentation (657)<input type="checkbox"/> Milestone (581)<input type="checkbox"/> Datamanagementplan (471)<input type="checkbox"/> Peerreview (459)<input type="checkbox"/> Patent (301)<input type="checkbox"/> Annotationcollection (255)	<input type="checkbox"/> Image (808195) - <ul style="list-style-type: none"><input type="checkbox"/> Figure (520873)<input type="checkbox"/> Photo (280895)<input type="checkbox"/> Drawing (4638)<input type="checkbox"/> Other (1176)<input type="checkbox"/> Plot (308)<input type="checkbox"/> Diagram (305) <input type="checkbox"/> Dataset (200178) <input type="checkbox"/> Software (106041) <input type="checkbox"/> Presentation (41341) <input type="checkbox"/> Other (20401) <input type="checkbox"/> Poster (17202) <input type="checkbox"/> Video (8283) <input type="checkbox"/> Lesson (4825) <input type="checkbox"/> Physicalobject (1127)

Các dạng dữ liệu nghiên cứu có thể được gắn DOI trên Zenodo



Cảm ơn!

<https://www.facebook.com/groups/OER.VN/>

oer-vn@googlegroups.com

Hỏi đáp

LÊ TRUNG NGHĨA

Email: nghialt@inoer.vn; letrungnghia.foss@gmail.com;

Blogs: <http://vnfoss.blogspot.com/>
<http://letrungnghia.mangvn.org/>